

SEQUENCE LISTING

**COPY**

(1) GENERAL INFORMATION:

- (i) APPLICANT: CRAIG, NANCY L
- (ii) TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN  
ATP-DEPENDENT TRANSPOSITION PROTEINS
- (iii) NUMBER OF SEQUENCES: 15
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Anne Brown (Alston & Bird, LLP)
  - (B) STREET: 3605 Glenwood Ave.
  - (C) CITY: Raleigh
  - (D) STATE: NC
  - (E) COUNTRY: USA
  - (F) ZIP: 27608
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Brown, Anne
  - (B) REGISTRATION NUMBER: 36,463
  - (C) REFERENCE/DOCKET NUMBER: 5789-3
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 919 420 2205
  - (B) TELEFAX: 919 881 3175

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1670 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..1668
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG AGT GCT ACC CGG ATT CAA GCA GTT TAT CGT GAT ACG GGG GTA GAG	48
Met Ser Ala Thr Arg Ile Gln Ala Val Tyr Arg Asp Thr Gly Val Glu	
1 5 10 15	
GCT TAT CGT GAT AAT CCT TTT ATC GAG GCC TTA CCA CCA TTA CAA GAG	96
Ala Tyr Arg Asp Asn Pro Phe Ile Glu Ala Leu Pro Pro Leu Gln Glu	
20 25 30	
TCA GTG AAT AGT GCT GCA TCA CTG AAA TCC TCT TTA CAG CTT ACT TCC	144
Ser Val Asn Ser Ala Ala Ser Leu Lys Ser Ser Leu Gln Leu Thr Ser	
35 40 45	
TCT GAC TTG CAA AAG TCC CGT GTT ATC AGA GCT CAT ACC ATT TGT CGT	192
Ser Asp Leu Gln Lys Ser Arg Val Ile Arg Ala His Thr Ile Cys Arg	
50 55 60	
ATT CCA GAT GAC TAT TTT CAG CCA TTA GGT ACG CAT TTG CTA CTA AGT	240
Ile Pro Asp Asp Tyr Phe Gln Pro Leu Gly Thr His Leu Leu Leu Ser	
65 70 75 80	
GAG CGT ATT TCG GTC ATG ATT CGA GGT GGC TAC GTA GGC AGA AAT CCT	288
Glu Arg Ile Ser Val Met Ile Arg Gly Gly Tyr Val Gly Arg Asn Pro	
85 90 95	
AAA ACA GGA GAT TTA CAA AAG CAT TTA CAA AAT GGT TAT GAG CGT GTT	336
Lys Thr Gly Asp Leu Gln Lys His Leu Gln Asn Gly Tyr Glu Arg Val	
100 105 110	
CAA ACG GGA GAG TTG GAG ACA TTT CGC TTT GAG GAG GCA CGA TCT ACG	384
Gln Thr Gly Glu Leu Glu Thr Phe Arg Phe Glu Glu Ala Arg Ser Thr	
115 120 125	
GCA CAA AGC TTA TTG TTA ATT GGT TGT TCT GGT AGT GGG AAG ACG ACC	432
Ala Gln Ser Leu Leu Leu Ile Gly Cys Ser Gly Ser Gly Lys Thr Thr	
130 135 140	
TCT CTT CAT CGT ATT CTA GCC ACG TAT CCT CAG GTG ATT TAC CAT CGT	480
Ser Leu His Arg Ile Leu Ala Thr Tyr Pro Gln Val Ile Tyr His Arg	
145 150 155 160	
GAA CTC AAT GTA GAG CAG GTG GTG TAT TTG AAA ATA GAC TGC TCG CAT	528
Glu Leu Asn Val Glu Gln Val Val Tyr Leu Lys Ile Asp Cys Ser His	
165 170 175	
AAT GGT TCG CTA AAA GAA ATC TGC TTG AAT TTT TTC AGA GCG TTG GAT	576
Asn Gly Ser Leu Lys Glu Ile Cys Leu Asn Phe Phe Arg Ala Leu Asp	
180 185 190	
CGA GCC TTG GGC TCG AAC TAT GAG CGT CGT TAT GGC TTA AAA CGT CAT	624
Arg Ala Leu Gly Ser Asn Tyr Glu Arg Arg Tyr Gly Leu Lys Arg His	
195 200 205	
GGT ATA GAA ACC ATG TTG GCT TTG ATG TCG CAA ATA GCC AAT GCA CAT	672
Gly Ile Glu Thr Met Leu Ala Leu Met Ser Gln Ile Ala Asn Ala His	
210 215 220	
GCT TTA GGG TTG TTG GTT ATT GAT GAA ATT CAG CAT TTA AGC CGC TCT	720

Ala	Leu	Gly	Leu	Leu	Val	Ile	Asp	Glu	Ile	Gln	His	Leu	Ser	Arg	Ser	
225					230					235						240
CGT	TCG	GGT	GGA	TCT	CAA	GAG	ATG	CTG	AAC	TTT	TTT	GTG	ACG	ATG	GTG	768
Arg	Ser	Gly	Gly	Ser	Gln	Glu	Met	Leu	Asn	Phe	Phe	Val	Thr	Met	Val	
				245					250					255		
AAT	ATT	ATT	GGC	GTA	CCA	GTG	ATG	TTG	ATT	GGT	ACC	CCT	AAA	GCA	CGA	816
Asn	Ile	Ile	Gly	Val	Pro	Val	Met	Leu	Ile	Gly	Thr	Pro	Lys	Ala	Arg	
			260					265					270			
GAG	ATT	TTT	GAG	GCT	GAT	TTG	CGG	TCT	GCA	CGT	AGA	GGG	GCA	GGG	TTT	864
Glu	Ile	Phe	Glu	Ala	Asp	Leu	Arg	Ser	Ala	Arg	Arg	Gly	Ala	Gly	Phe	
		275					280					285				
GGA	GCT	ATA	TTC	TGG	GAT	CCT	ATA	CAA	CAA	ACG	CAA	CGT	GGA	AAG	CCC	912
Gly	Ala	Ile	Phe	Trp	Asp	Pro	Ile	Gln	Gln	Thr	Gln	Arg	Gly	Lys	Pro	
		290				295					300					
AAT	CAA	GAG	TGG	ATC	GCT	TTT	ACG	GAT	AAT	CTT	TGG	CAA	TTA	CAG	CTT	960
Asn	Gln	Glu	Trp	Ile	Ala	Phe	Thr	Asp	Asn	Leu	Trp	Gln	Leu	Gln	Leu	
305				310						315				320		
TTA	CAA	CGC	AAA	GAT	GCG	CTG	TTA	TCG	GAT	GAG	GTC	CGT	GAT	GTG	TGG	1008
Leu	Gln	Arg	Lys	Asp	Ala	Leu	Leu	Ser	Asp	Glu	Val	Arg	Asp	Val	Trp	
			325						330					335		
TAT	GAG	CTA	AGC	CAA	GGA	GTG	ATG	GAC	ATT	GTA	GTA	AAA	CTT	TTT	GTA	1056
Tyr	Glu	Leu	Ser	Gln	Gly	Val	Met	Asp	Ile	Val	Val	Lys	Leu	Phe	Val	
			340					345					350			
CTC	GCT	CAG	CTC	CGT	GCG	CTA	GCT	TTA	GGC	AAT	GAG	CGT	ATT	ACC	GCT	1104
Leu	Ala	Gln	Leu	Arg	Ala	Leu	Ala	Leu	Gly	Asn	Glu	Arg	Ile	Thr	Ala	
		355					360					365				
GGT	TTA	TTG	CGG	CAA	GTG	TAT	CAA	GAT	GAG	TTA	AAG	CCT	GTG	CAC	CCC	1152
Gly	Leu	Leu	Arg	Gln	Val	Tyr	Gln	Asp	Glu	Leu	Lys	Pro	Val	His	Pro	
	370					375					380					
ATG	CTA	GAG	GCA	TTA	CGC	TCG	GGT	ATC	CCA	GAA	CGC	ATT	GCT	CGT	TAT	1200
Met	Leu	Glu	Ala	Leu	Arg	Ser	Gly	Ile	Pro	Glu	Arg	Ile	Ala	Arg	Tyr	
385					390					395				400		
TCT	GAT	CTA	GTC	GTT	CCC	GAG	ATT	GAT	AAA	CGG	TTA	ATC	CAA	CTT	CAG	1248
Ser	Asp	Leu	Val	Val	Pro	Glu	Ile	Asp	Lys	Arg	Leu	Ile	Gln	Leu	Gln	
				405					410					415		
CTA	GAT	ATC	GCA	GCG	ATA	CAA	GAA	CAA	ACA	CCA	GAA	GAA	AAA	GCC	CTT	1296
Leu	Asp	Ile	Ala	Ala	Ile	Gln	Glu	Gln	Thr	Pro	Glu	Glu	Lys	Ala	Leu	
			420					425					430			
CAA	GAG	TTA	GAT	ACC	GAA	GAT	CAG	CGT	CAT	TTA	TAT	CTG	ATG	CTG	AAA	1344
Gln	Glu	Leu	Asp	Thr	Glu	Asp	Gln	Arg	His	Leu	Tyr	Leu	Met	Leu	Lys	
		435					440					445				
GAG	GAT	TAC	GAT	TCA	AGC	CTG	TTA	ATT	CCC	ACT	ATT	AAA	AAA	GCG	TTT	1392
Glu	Asp	Tyr	Asp	Ser	Ser	Leu	Leu	Ile	Pro	Thr	Ile	Lys	Lys	Ala	Phe	
	450					455					460					

AGC	CAG	AAT	CCA	ACG	ATG	ACA	AGA	CAA	AAG	TTA	CTG	CCT	CTT	GTT	TTG	1440
Ser	Gln	Asn	Pro	Thr	Met	Thr	Arg	Gln	Lys	Leu	Leu	Pro	Leu	Val	Leu	
465					470				475						480	
CAG	TGG	TTG	ATG	GAA	GGC	GAA	ACG	GTA	GTG	TCA	GAA	CTA	GAA	AAG	CCC	1488
Gln	Trp	Leu	Met	Glu	Gly	Glu	Thr	Val	Val	Ser	Glu	Leu	Glu	Lys	Pro	
				485					490					495		
TCC	AAG	AGT	AAA	AAG	GTT	TCG	GCT	ATA	AAG	GTA	GTC	AAG	CCC	AGC	GAC	1536
Ser	Lys	Ser	Lys	Lys	Val	Ser	Ala	Ile	Lys	Val	Val	Lys	Pro	Ser	Asp	
			500					505					510			
TGG	GAT	AGC	TTG	CCT	GAT	ACG	GAT	TTA	CGT	TAT	ATC	TAT	TCA	CAA	CGC	1584
Trp	Asp	Ser	Leu	Pro	Asp	Thr	Asp	Leu	Arg	Tyr	Ile	Tyr	Ser	Gln	Arg	
	515						520					525				
CAA	CCT	GAA	AAA	ACC	ATG	CAT	GAA	CGG	TTA	AAA	GGG	AAA	GGG	GTA	ATA	1632
Gln	Pro	Glu	Lys	Thr	Met	His	Glu	Arg	Leu	Lys	Gly	Lys	Gly	Val	Ile	
	530					535					540					
GTG	GAT	ATG	GCG	AGC	TTA	TTT	AAA	CAA	GCA	GGT	TAG	CC				1670
Val	Asp	Met	Ala	Ser	Leu	Phe	Lys	Gln	Ala	Gly	*					
545					550				555							

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 556 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Ser	Ala	Thr	Arg	Ile	Gln	Ala	Val	Tyr	Arg	Asp	Thr	Gly	Val	Glu	
1				5					10					15		
Ala	Tyr	Arg	Asp	Asn	Pro	Phe	Ile	Glu	Ala	Leu	Pro	Pro	Leu	Gln	Glu	
			20					25					30			
Ser	Val	Asn	Ser	Ala	Ala	Ser	Leu	Lys	Ser	Ser	Leu	Gln	Leu	Thr	Ser	
		35					40					45				
Ser	Asp	Leu	Gln	Lys	Ser	Arg	Val	Ile	Arg	Ala	His	Thr	Ile	Cys	Arg	
	50					55					60					
Ile	Pro	Asp	Asp	Tyr	Phe	Gln	Pro	Leu	Gly	Thr	His	Leu	Leu	Leu	Ser	
65					70					75				80		
Glu	Arg	Ile	Ser	Val	Met	Ile	Arg	Gly	Gly	Tyr	Val	Gly	Arg	Asn	Pro	
			85					90						95		
Lys	Thr	Gly	Asp	Leu	Gln	Lys	His	Leu	Gln	Asn	Gly	Tyr	Glu	Arg	Val	
			100					105					110			

Gln Thr Gly Glu Leu Glu Thr Phe Arg Phe Glu Glu Ala Arg Ser Thr  
 115 120 125  
 Ala Gln Ser Leu Leu Leu Ile Gly Cys Ser Gly Ser Gly Lys Thr Thr  
 130 135 140  
 Ser Leu His Arg Ile Leu Ala Thr Tyr Pro Gln Val Ile Tyr His Arg  
 145 150 155 160  
 Glu Leu Asn Val Glu Gln Val Val Tyr Leu Lys Ile Asp Cys Ser His  
 165 170 175  
 Asn Gly Ser Leu Lys Glu Ile Cys Leu Asn Phe Phe Arg Ala Leu Asp  
 180 185 190  
 Arg Ala Leu Gly Ser Asn Tyr Glu Arg Arg Tyr Gly Leu Lys Arg His  
 195 200 205  
 Gly Ile Glu Thr Met Leu Ala Leu Met Ser Gln Ile Ala Asn Ala His  
 210 215 220  
 Ala Leu Gly Leu Leu Val Ile Asp Glu Ile Gln His Leu Ser Arg Ser  
 225 230 235 240  
 Arg Ser Gly Gly Ser Gln Glu Met Leu Asn Phe Phe Val Thr Met Val  
 245 250 255  
 Asn Ile Ile Gly Val Pro Val Met Leu Ile Gly Thr Pro Lys Ala Arg  
 260 265 270  
 Glu Ile Phe Glu Ala Asp Leu Arg Ser Ala Arg Arg Gly Ala Gly Phe  
 275 280 285  
 Gly Ala Ile Phe Trp Asp Pro Ile Gln Gln Thr Gln Arg Gly Lys Pro  
 290 295 300  
 Asn Gln Glu Trp Ile Ala Phe Thr Asp Asn Leu Trp Gln Leu Gln Leu  
 305 310 315 320  
 Leu Gln Arg Lys Asp Ala Leu Leu Ser Asp Glu Val Arg Asp Val Trp  
 325 330 335  
 Tyr Glu Leu Ser Gln Gly Val Met Asp Ile Val Val Lys Leu Phe Val  
 340 345 350  
 Leu Ala Gln Leu Arg Ala Leu Ala Leu Gly Asn Glu Arg Ile Thr Ala  
 355 360 365  
 Gly Leu Leu Arg Gln Val Tyr Gln Asp Glu Leu Lys Pro Val His Pro  
 370 375 380  
 Met Leu Glu Ala Leu Arg Ser Gly Ile Pro Glu Arg Ile Ala Arg Tyr  
 385 390 395 400  
 Ser Asp Leu Val Val Pro Glu Ile Asp Lys Arg Leu Ile Gln Leu Gln  
 405 410 415  
 Leu Asp Ile Ala Ala Ile Gln Glu Gln Thr Pro Glu Glu Lys Ala Leu

420	425	430
Gln Glu Leu Asp Thr Glu Asp	Gln Arg His Leu Tyr	Leu Met Leu Lys
435	440	445
Glu Asp Tyr Asp Ser Ser Leu	Leu Ile Pro Thr	Ile Lys Lys Ala Phe
450	455	460
Ser Gln Asn Pro Thr Met Thr	Arg Gln Lys Leu	Leu Pro Leu Val Leu
465	470	475 480
Gln Trp Leu Met Glu Gly Glu	Thr Val Val Ser Glu	Leu Glu Lys Pro
485	490	495
Ser Lys Ser Lys Lys Val Ser	Ala Ile Lys Val	Val Lys Pro Ser Asp
500	505	510
Trp Asp Ser Leu Pro Asp Thr	Asp Leu Arg Tyr	Ile Tyr Ser Gln Arg
515	520	525
Gln Pro Glu Lys Thr Met His	Glu Arg Leu Lys	Gly Lys Gly Val Ile
530	535	540
Val Asp Met Ala Ser Leu Phe	Lys Gln Ala Gly	*
545	550	555

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5926 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "pEM delta R.adj to 1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTTAGAGCAA TTCGGTGTTA GTTTCAGCAA GCAAACATTA ACCATAGCTA ATGATTTATA	60
GCCATATTAA CCATTGGGGT ACCGAGCTCG AATTCCATGG TCTGTTTCCT GTGTGAAATT	120
GTTATCCGCT CACAATTCCA CACATTATAC GAGCCGGATG ATTAATTGTC AACAGCTCAT	180
TTCAGAATAT TTGCCAGAAC CGTTATGATG TCGGCGCAAA AAACATTATC CAGAACGGGA	240
GTGCGCCTTG AGCGACACGA ATTATGCAGT GATTTACGAC CTGCACAGCC ATACCACAGC	300
TTCCGATGGC TGCCTGACGC CAGAAGCATT GGTGCACCGT GCAGTCGATG ATAAGCTGTC	360
AAACCAGATC AATTGCGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA	420
GTCGGGAAAC CTGTCGTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG	480

TTTGCGTATT GGGCGCCAGG GTGGTTTTTC TTTTCACCAG TGAGACGGGC AACAGCTGAT 540  
 TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT GCAGCAAGCG GTCCACGCTG GTTTGCCCCA 600  
 GCAGGCGAAA ATCCTGTTTG ATGGTGGTTG ACGGCGGGAT ATAACATGAG CTGTCTTCGG 660  
 TATCGTCGTA TCCCACTACC GAGATATCCG CACCAACGCG CAGCCCGGAC TCGGTAATGG 720  
 CGCGCATTGC GCCCAGCGCC ATCTGATCGT TGGCAACCAG CATCGCAGTG GGAACGATGC 780  
 CCTCATTCAG CATTTGCATG GTTTGTTGAA AACC GGACAT GGC ACTCCAG TCGCCTTCCC 840  
 GTTCCGCTAT CGGCTGAATT TGATTGCGAG TGAGATATTT ATGCCAGCCA GCCAGACGCA 900  
 GACGCGCCGA GACAGAACTT AATGGGCCCC CTAACAGCGC GATTTGCTGG TGACCCAATG 960  
 CGACCAGATG CTCCACGCCC AGTCGCGTAC CGTCTTCATG GGAGAAAATA ATACTGTTGA 1020  
 TGGGTGTCTG GTCAGAGACA TCAAGAAATA ACGCCGGAAC ATTAGTGAGC GCAGCTTCCA 1080  
 CAGCAATGGC ATCCTGGTCA TCCAGCGGAT AGTTAATGAT CAGCCCCTG ACGCGTTGCG 1140  
 CGAGAAGATT GTGCACCGCC GCTTTACAGG CTTTCGACGCC GCTTCGTTCT ACCATCGACA 1200  
 CCACCACGCT GGCACCCAGT TGATCGGCGC GAGATTTAAT CGCCGCGACA ATTTGCGACG 1260  
 GCGCGTGCAG GGCCAGACTG GAGGTGGCAA CGCCAATCAG CAACGACTGT TTGCCCCCCA 1320  
 GTTGTGTGTC CACGCGGTTG GGAATGTAAT TCAGCTCCGC CATCGCCGCT TCCACTTTTT 1380  
 CCCGCGTTTT CGCAGAAACG TGGCTGGCCT GGTTCACCAC GCGGGAAACG GTCTGATAAG 1440  
 AGACACCGGC ATACTCTGCG ACATCGTATA ACGTTACTGG TTTACATTC ACCACCCTGA 1500  
 ATTGACTCTC TTCCGGGCGC TATCATGCCA TACCGCGAAA GGTTTTGCAC CATTGATGG 1560  
 TGTCAACGTA AATGCATGCC GCTTCGCCTT CGCGCGCGAA TTGATCTGCT GCCTCGCGCG 1620  
 TTTGCGTGAT GACGGTGAAA ACCTCTGACA CATGCAGCTC CCGGAGACGG TCACAGCTTG 1680  
 TCTGTAAGCG GATGCCGGGA GCAGACAAGC CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG 1740  
 GTGTCGGGGC GCAGCCATGA CCCAGTCACG TAGCGATAGC GGAGTGTATA CTGGCTTAAC 1800  
 TATGCGGCAT CAGAGCAGAT TGTACTGAGA GTGCACCATA TGCGGTGTGA AATACCGCAC 1860  
 AGATGCGTAA GGAGAAAATA CCGCATCAGG CGCTCTTCCG CTTCTCGCT CACTGACTCG 1920  
 CTGCGCTCGG TCGTTGCGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG 1980  
 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG 2040  
 GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC 2100  
 GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA 2160  
 TACCAGGCGT TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT 2220

ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA TAGCTCACGC 2280  
 TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC 2340  
 CCCGTTTCAGC CCGACCGCTG CGCCTTATCC GGTAACATATC GTCTTGAGTC CAACCCGGTA 2400  
 AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT 2460  
 GTAGGCGGTG CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGGACA 2520  
 GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT 2580  
 TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT 2640  
 ACGCGCAGAA AAAAAGGATC TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT 2700  
 CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC 2760  
 ACCTAGATCC TTTTAAATTA AAAATGAAGT TTAAATCAA TCTAAAGTAT ATATGAGTAA 2820  
 ACTTGGTCTG ACAGTTACCA ATGCTTAATC AGTGAGGCAC CTATCTCAGC GATCTGTCTA 2880  
 TTTTCGTTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC 2940  
 TTACCATCTG GCCCCAGTGC TGCAATGATA CCGCGAGACC CACGCTCACC GGCTCCAGAT 3000  
 TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA 3060  
 TCCGCCTCCA TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT 3120  
 AATAGTTTGC GCAACGTTGT TGCCATTGCT GTAGGCATCG TGGTGTACAG CTCGTCGTTT 3180  
 GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCCATG 3240  
 TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT CCTCCGATCG TTGTCAGAAG TAAGTTGGCC 3300  
 GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC 3360  
 GTAAGATGCT TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG 3420  
 CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ACACGGGATA ATACCGCGCC ACATAGCAGA 3480  
 ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC AAGGATCTTA 3540  
 CCGCTGTTGA GATCCAGTTC GATGTAACCC ACTCGTGCAC CCAACTGATC TTCAGCATCT 3600  
 TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAG 3660  
 GGAATAAGGG CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA 3720  
 AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT 3780  
 AAACAAAAG AGTTTGTAGA AACGCAAAAA GGCCATCCGT CAGGATGGCC TTCTGCTTAA 3840  
 TTTGATGCCT GGCAGTTTAT GCGGGCGTC CTGCCCCCA CCCTCCGGGC CGTTGCTTCG 3900  
 CAACGTTCAA ATCCGCTCCC GCGGATTG TCCTACTCAG GAGAGCGTTC ACCGACAAAC 3960



AACAGATAAA ACGAAAGGCC CAGTCTTTCG ACTGAGCCTT TCGTTTTATT TGATGCCTGG 4020  
CAGTTCCCTA CTCTCGCATG GGGAGACCCC ACACTACCAT CGGCGCTACG GCGTTTCACT 4080  
TCTGAGTTCG GCATGGGGTC AGGTGGGACC ACCGCGCTAC TGCCGCCAGG CAAATTCTGT 4140  
TTTATCAGAC CGCTTCTGCG TTCTGATTTA ATCTGTATCA GGCTGAAAAT CTTCTCTCAT 4200  
CCGCCAAAAC AGCCAAGCTT GCATGCCTGC AGGTCGACTC TAGAGGATCC CCAAGAAAGT 4260  
CCGTCGGACA GCTTTAATAA ACCCTGCACT TATCTGTTTA GTGTGGGCGG ACAAATAGT 4320  
TGGGAAGTGG GAGGGGTGGA AATGGAGTTT TTAAGGATTA TTTAGGGAAG AGTGACAAAA 4380  
TAGATGGGAA CTGGGTGTAG CGTCGTAAGC TAATACGAAA ATTAAAAATG ACAAATAGT 4440  
TTGGAAGTAG ATTTCACTTA TCTGGTTGGT CGACCTGCAG GGGGGGGGGG GAAAGCCACG 4500  
TTGTGTCTCA AAATCTCTGA TGTTACATTG CACAAGATAA AAATATATCA TCATGAACAA 4560  
TAAACTGTC TGCTTACATA AACAGTAATA CAAGGGGTGT TATGAGCCAT ATTCAACGGG 4620  
AAACGTCTTG CTCGAGGCCG CGATTAAATT CCAACATGGA TGCTGATTTA TATGGGTATA 4680  
AATGGGCTCG CGATAATGTC GGGCAATCAG GTGCGACAAT CTATCGATTG TATGGGAAGC 4740  
CCGATGCGCC AGAGTTGTTT CTGAAACATG GCAAAGGTAG CGTTGCCAAT GATGTTACAG 4800  
ATGAGATGGT CAGACTAAAC TGGCTGACGG AATTTATGCC TCTTCCGACC ATCAAGCATT 4860  
TTATCCGTAC TCCTGATGAT GCATGGTTAC TCACCACTGC GATCCCCGGG AAAACAGCAT 4920  
TCCAGGTATT AGAAGAATAT CCTGATTCAG GTGAAAATAT TGTTGATGCG CTGGCAGTGT 4980  
TCCTGCGCCG GTTGCAATCG ATTCTGTTT GTAATTGTCC TTTTAACAGC GATCGCGTAT 5040  
TTCGTCTCGC TCAGGCGCAA TCACGAATGA ATAACGGTTT GGTGATGCG AGTGATTTTG 5100  
ATGACGAGCG TAATGGCTGG CCTGTTGAAC AAGTCTGGAA AGAAATGCAT AAGCTTTTGC 5160  
CATTCTCACC GGATTTCAGT GTCACATG GTGATTTCTC ACTTGATAAC CTTATTTTTG 5220  
ACGAGGGGAA ATTAATAGGT TGTATTGATG TTGGACGAGT CGGAATCGCA GACCGATACC 5280  
AGGATCTTGC CATCCTATGG AACTGCCTCG GTGAGTTTTT TCCTTCATTA CAGAAACGGC 5340  
TTTTTCAAAA ATATGGTATT GATAATCCTG ATATGAATAA ATTGCAGTTT CATTTGATGC 5400  
TCGATGAGTT TTTCTAATCA GAATTGGTTA ATTGGTTGTA ACACTGGCAG AGCATTACGC 5460  
TGACTTGACG GGACGGCGGC TTTGTTGAAT AAATCGAACT TTTGCTGAGT TGAAGGATCA 5520  
GATCACGCAT CTTCCCGACA ACGCAGACCG TTCCGTGGCA AAGCAAAAGT TCAAAATCAC 5580  
CAACTGGTCC ACCTACAACA AAGCTCTCAT CAACCGTGGC TCCCTCACTT TCTGGCTGGA 5640  
TGATGGGGCG ATTCAGGCCT GGTATGAGTC AGCAACACCT TCTTCACGAG GCAGACCTCA 5700

GCGCCCCCCC CCCCCTGCAG GTCGACCCCA CGCCCCTCTT TAATACGACG GGCAATTTGC 5760  
 ACTTCAGAAA ATGAAGAGTT TGCTTTAGCC ATAACAAAAG TCCAGTATGC TTTTTCACAG 5820  
 CATAACTGGA CTGATTTTCAG TTTACAAC TA TCTGTCTAG TTTAAGACTT TATTGTCATA 5880  
 GTTTAGATCT ATTTTGTTC GTTTAAGACT TTATTGTCCG CCCACA 5926

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5926 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "pEM-delta"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CAGATCAATT CGCGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCCG TTTCCAGTCG 60  
 GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG 120  
 CGTATTGGGC GCCAGGGTGG TTTTCTTTT CACCAGTGAG ACGGGCAACA GCTGATTGCC 180  
 CTTACCGCC TGGCCCTGAG AGAGTTGCAG CAAGCGGTCC ACGCTGGTTT GCCCCAGCAG 240  
 GCGAAAATCC TGTTTGATGG TGTTGACGG CGGGATATAA CATGAGCTGT CTTCGGTATC 300  
 GTCGTATCCC ACTACCGAGA TATCCGCACC AACGCGCAGC CCGGACTCGG TAATGGCGCG 360  
 CATTGCGCCC AGCGCCATCT GATCGTTGGC AACCAGCATC GCAGTGGGAA CGATGCCCTC 420  
 ATTCAGCATT TGCATGGTTT GTTGAAAACC GGACATGGCA CTCCAGTCGC CTTCCCGTTC 480  
 CGCTATCGGC TGAATTTGAT TGCGAGTGAG ATATTTATGC CAGCCAGCCA GACGCAGACG 540  
 CGCCGAGACA GAACTTAATG GGCCCGCTAA CAGCGCGATT TGCTGGTGAC CCAATGCGAC 600  
 CAGATGCTCC ACGCCCAGTC GCGTACCGTC TTCATGGGAG AAAATAATAC TGTTGATGGG 660  
 TGTCTGGTCA GAGACATCAA GAAATAACGC CGGAACATTA GTGCAGGCAG CTTCCACAGC 720  
 AATGGCATCC TGGTCATCCA GCGGATAGTT AATGATCAGC CCACTGACGC GTTGCGCGAG 780  
 AAGATTGTGC ACCGCCGCTT TACAGGCTTC GACGCCGCTT CGTTCTACCA TCGACACCAC 840  
 CACGCTGGCA CCCAGTTGAT CGGCGCGAGA TTTAATCGCC GCGACAATTT GCGACGGCGC 900  
 GTGCAGGGCC AGACTGGAGG TGGCAACGCC AATCAGCAAC GACTGTTTGC CCGCCAGTTG 960  
 TTGTGCCACG CGGTTGGGAA TGTAATTCAG CTCCGCCATC GCCGCTTCCA CTTTTTCCCG 1020

CGTTTTTCGCA GAAACGTGGC TGGCCTGGTT CACCACGCGG GAAACGGTCT GATAAGAGAC 1080

ACCGGCATAC TCTGCGACAT CGTATAACGT TACTGGTTTC ACATTCACCA CCCTGAATTG 1140

ACTCTCTTCC GGGCGCTATC ATGCCATACC GCGAAAGGTT TTGCACCATT CGATGGTGTC 1200

AACGTAAATG CATGCCGCTT CGCCTTCGCG CGCGAATTGA TCTGCTGCCT CGCGCGTTTC 1260

GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCGG AGACGGTCAC AGCTTGCTG 1320

TAAGCGGATG CCGGGAGCAG ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT 1380

CGGGGCGCAG CCATGACCCA GTCACGTAGC GATAGCGGAG TGTATACTGG CTTAACATG 1440

CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT 1500

GCGTAAGGAG AAAATACCGC ATCAGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC 1560

GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT 1620

CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA 1680

GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC 1740

ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC 1800

AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG 1860

GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 1920

GGTATCTCAG TTCGGTGTAG GTCGTTGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG 1980

TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC 2040

ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG 2100

GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACACTAG CTACACTAGA AGGACAGTAT 2160

TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT 2220

CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTTGT TTGCAAGCAG CAGATTACGC 2280

GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT 2340

GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT 2400

AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT 2460

GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTT 2520

GTTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAAGATAAC TACGATACGG GAGGGCTTAC 2580

CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT 2640

CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA ACTTTATCCG 2700

CCTCCATCCA GTCTATTAAT TGTGCGGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA 2760

GTTTGCGCAA CGTTGTTGCC ATTGCTGTAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA 2820  
 TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC CCCATGTTGT 2880  
 GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG 2940  
 TGTTATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA 3000  
 GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC 3060  
 GACCGAGTTG CTCTTGCCCG GCGTCAACAC GGGATAATAC CGCGCCACAT AGCAGAACTT 3120  
 TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC 3180  
 TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA 3240  
 CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA 3300  
 TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA 3360  
 TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 3420  
 AAAAAGAGTT TGTAGAAACG CAAAAAGGCC ATCCGTCAGG ATGGCCTTCT GCTTAATTTG 3480  
 ATGCCTGGCA GTTTATGGCG GCGTCCTGC CCGCCACCCT CCGGGCCGTT GCTTCGCAAC 3540  
 GTTCAAATCC GCTCCCGGCG GATTTGTCCT ACTCAGGAGA GCGTTCACCG ACAAACAACA 3600  
 GATAAAACGA AAGGCCCAGT CTTTCGACTG AGCCTTTCGT TTTATTTGAT GCCTGGCAGT 3660  
 TCCCTACTCT CGCATGGGGA GACCCACAC TACCATCGGC GCTACGGCGT TTCACTTCTG 3720  
 AGTTCGGCAT GGGGTCAGGT GGGACCACCG CGCTACTGCC GCCAGGCAAA TTCTGTTTTA 3780  
 TCAGACCGCT TCTGCGTTCT GATTTAATCT GTATCAGGCT GAAAATCTTC TCTCATCCGC 3840  
 CAAAACAGCC AAGCTTGCAT GCCTGCAGGT CGACTCTAGA GGATCCCCAA GAAAGTCCGT 3900  
 CGGACAGCTT TAATAAACCC TGCACCTATC TGTTTAGTGT GGGCGGACAA AATAGTTGGG 3960  
 AACTGGGAGG GGTGGAAATG GAGTTTTTAA GGATTATTTA GGGAAGAGTG ACAAATAGA 4020  
 TGGGAACTGG GTGTAGCGTC GTAAGCTAAT ACGAAAATTA AAAATGACAA AATAGTTTGG 4080  
 AACTAGATTT CACTTATCTG GTTGGTCGAC CTGCAGGGGG GGGGGGGAAA GCCACGTTGT 4140  
 GTCTCAAAAT CTCTGATGTT ACATTGCACA AGATAAAAAT ATATCATCAT GAACAATAAA 4200  
 ACTGTCTGCT TACATAAACA GTAATACAAG GGGTGTTATG AGCCATATTC AACGGGAAAC 4260  
 GTCTTGCTCG AGGCCGCGAT TAAATTCCAA CATGGATGCT GATTTATATG GGTATAAATG 4320  
 GGCTCGCGAT AATGTCGGGC AATCAGGTGC GACAATCTAT CGATTGTATG GGAAGCCCGA 4380  
 TCGCCAGAG TTGTTTCTGA AACATGGCAA AGGTAGCGTT GCCAATGATG TTACAGATGA 4440  
 GATGGTCAGA CTAACTGGC TGACGGAATT TATGCCTCTT CCGACCATCA AGCATTTTAT 4500

```

CCGTACTCCT GATGATGCAT GGTACTCAC CACTGCGATC CCCGGGAAAA CAGCATTCCA 4560
GGTATTAGAA GAATATCCTG ATTCAGGTGA AAATATTGTT GATGCGCTGG CAGTGTTCCT 4620
GCGCCGGTTG CATTGATTTC CTGTTTGTA TGTCTCTTTT AACAGCGATC GCGTATTTTCG 4680
TCTCGCTCAG GCGCAATCAC GAATGAATAA CGGTTTGGTT GATGCGAGTG ATTTTGATGA 4740
CGAGCGTAAT GGCTGGCCTG TTGAACAAGT CTGAAAAGAA ATGCATAAGC TTTTGCCATT 4800
CTCACCGGAT TCAGTCGTCA CTCATGGTGA TTTCTCACTT GATAACCTTA TTTTGTACGA 4860
GGGGAAATTA ATAGGTTGTA TTGATGTTGG ACGAGTCGGA ATCGCAGACC GATACCAGGA 4920
TCTTGCCATC CTATGGAAC TCCCTCGGTGA GTTTTCTCCT TCATTACAGA AACGGCTTTT 4980
TCAAAAATAT GGTATTGATA ATCCTGATAT GAATAAATTG CAGTTTCATT TGATGCTCGA 5040
TGAGTTTTTC TAATCAGAAT TGGTTAATTG GTTGTAACAC TGGCAGAGCA TTACGCTGAC 5100
TTGACGGGAC GCGGGCTTTG TTGAATAAAT CGAACTTTTG CTGAGTTGAA GGATCAGATC 5160
ACGCATCTTC CCGACAACGC AGACCGTTCC GTGGCAAAGC AAAAGTTCAA AATCACCAAC 5220
TGGTCCACCT ACAACAAAGC TCTCATCAAC CGTGGCTCCC TCACTTTCTG GCTGGATGAT 5280
GGGGCGATT CAGGCTGGTA TGAGTCAGCA ACACCTTCTT CACGAGGCAG ACCTCAGCGC 5340
CCCCCCCCCT CTGAGGTCG ACCCCACGCC CCTCTTTAAT ACGACGGGCA ATTTGCACTT 5400
CAGAAAATGA AGAGTTTGCT TTAGCCATAA CAAAAGTCCA GTATGCTTTT TCACAGCATA 5460
ACTGGACTGA TTTCACTTTA CAACTATTCT GTCTAGTTTA AGACTTTATT GTCATAGTTT 5520
AGATCTATTT TGTTCAGTTT AAGACTTTAT TGTCCGCCCA CATTTAGAGC AATTCGGTGT 5580
TAGTTTCAGC AAGCAAACAT TAACCATAGC TAATGATTTA TAGCCATATT AACCATTGGG 5640
GTACCGAGCT CGAATTCCAT GGTCTGTTTC CTGTGTGAAA TTGTTATCCG CTCACAATTC 5700
CACACAATAT ACGAGCCGGA TGATTAATTG TCAACAGCTC ATTTTCAGAAT ATTTGCCAGA 5760
ACCGTTATGA TGTGCGCGCA AAAAACATTA TCCAGAACGG GAGTGCGCCT TGAGCGACAC 5820
GAATTATGCA GTGATTTACG ACCTGCACAG CCATACCACA GCTTCCGATG GCTGCCTGAC 5880
GCCAGAAGCA TTGGTGCACC GTGCAGTCGA TGATAAGCTG TCAAAC 5926

```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "pER183 (target plasmid)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAATTCCGGA TGAGCATTCA TCAGGCGGGC AAGAATGTGA ATAAAGGCCG GATAAACTT 60  
GTGCTTATTT TTCTTTACGG TCTTTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT 120  
ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAATGT TCTTTACGAT GCCATTGGGA 180  
TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA 240  
AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTTATTAT GGTGAAAGTT 300  
GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC 360  
GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG TCACAGGTAT 420  
TTATTCGGCG CAAAGTGC GT GCGGTGATGC TGCCAACTTA CTGATTTAGT GTATGATGGT 480  
GTTTTTTGAGG TGCTCCAGTG GCTTCTGTTT CTATCAGCTG TCCCTCCTGT TCAGCTACTG 540  
ACGGGGTGGT GCGTAACGGC AAAAGCACCG CCGGACATCA GCGCTAGCGG AGTGTATACT 600  
GGCTTACTAT GTTGGCACTG ATGAGGGTGT CAGTGAAGTG CTTCATGTGG CAGGAGAAAA 660  
AAGGCTGCAC CGGTGCGTCA GCAGAATATG TGATACAGGA TATATTCCGC TTCCTCGCTC 720  
ACTGACTCGC TACGCTCGGT CGTTCGACTG CGGCGAGCGG AAATGGCTTA CGAACGGGGC 780  
GGAGATTTCC TGGAAGATGC CAGGAAGATA CTTAACAGGG AAGTGAGAGG GCCGCGGCAA 840  
AGCCGTTTTT CCATAGGCTC CGCCCCCTG ACAAGCATCA CGAAATCTGA CGCTCAAATC 900  
AGTGGTGGCG AAACCCGACA GGAATATAAA GATACCAGGC GTTTCCCCTG GCGGCTCCCT 960  
CGTGCGCTCT CCTGTTCTG CCTTTCGGTT TACCGGTGTC ATTCCGCTGT TATGGCCGCG 1020  
TTTGTCTCAT TCCACGCCTG AACTCAGTT CCGGGTAGGC AGTTCGCTCC AAGCTGGACT 1080  
GTATGCACGA ACCCCCCGTT CAGTCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG 1140  
AGTCCAACCC GGAAAGACAT GCAAAAGCAC CACTGGCAGC AGCCACTGGT AATTGATTTA 1200  
GAGGAGTTAG TCTTGAAGTC ATGCGCCGGT TAAGGCTAAA CTGAAAGGAC AAGTTTTGGT 1260  
GACTGCGCTC CTCCAAGCCA GTTACCTCGG TTCAAAGAGT TGGTAGCTCA GAGAACCTTC 1320  
GAAAAACCGC CCTGCAAGGC GGTTTTTTCG TTTTCAGAGC AAGAGATTAC GCGCAGACCA 1380  
AAACGATCTC AAGAAGATCA TCTTATTAAT CAGATAAAAT ATTTCTAGAT TTCAGTGCAA 1440  
TTTATCTCTT CAAATGTAGC ACCTGAAGTC AGCCCCATAC GATATAAGTT GTAATTCTCA 1500  
TGTTTGACAG CTTATCATCG GATGGATCTG AAATTGTAAA CGTTAATATT TTGTTAAATT 1560

CGCGTTAAAT TTTTGTAAAA TCAGCTCATT TTTTAACCAA TAGGCCGAAA TCGGCAAAAT 1620

CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGGTTGAGT GTTGTTCAG TTTGGAACAA 1680

GAGTCCACTA TTAAAGAACG TGGACTCCAA CGTCAAAGGG CGAAAAACCG TCTATCAGGG 1740

CGATGGCCCA CTACGTGAAC CATCACCTA ATCAAGTTTT TTGGGGTCGA GGTGCCGTAA 1800

AGCACTAAAT CGGAACCCTA AAGGGAGCCC CCGATTAGA GCTTGACGGG GAAAGCCGGC 1860

GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG GGCGCTAGGG CGCTGGCAAG 1920

TGTAGCGGTC ACGCTGCGCG TAACCACCAC ACCCGCCGCG CTTAATGCGC CGCTACAGGG 1980

CGCGTCAGAT CCCATCGATA AGCTTTAATG CGGTAGTTTA TCACAGTTAA ATTGCTAACG 2040

CAGTCAGGCA CCGTGTATGA AATCTAACAA TGCCTCATC GTCATCCTCG GCACCGTCAC 2100

CCTGGATGCT GTAGGCATAG GCTTGGTTAT GCCGGTACTG CCGGGCCTCT TCGGGGATAT 2160

CGTCCATTCC GACAGCATCG CCAGTCACTA TGGCGTGCTG CTAGCGCTAT ATGCGTTGAT 2220

GCAATTTCTA TGCGCACCCG TTCTCGGAGC ACTGTCCGAC CGCTTTGGCC GCCGCCCAGT 2280

CCTGCTCGCT TCGCTACTTG GAGCCACTAT CGACTACGCG ATCATGGCGA CCACACCCGT 2340

CCTGTGGATC CGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA 2400

GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTGCGGCC GCCCTGCAAG 2460

GAAGGGAATG TCGCCAACAG CGAAGAGAGT TGGGCAACGG ATGTGCTGGT GGAGGTGATC 2520

GCCTCCTGAT GATGAGCCGC TCCCGATGTG GTGTCGGGAG CGGTATTTTC TATAAACTT 2580

ACCGCTTATT TGAGATATTC ATCGAAAATG TCGAGTAATT CTTGATGTAT ACACGGCCAT 2640

TCCTGACCTA AATTGACGGT ACACAAGCCA ATATCGAAGC CATTAATTTT ATAACGATGT 2700

TTCCTGCGG TATCTACGTG GGGATATATT AATAACCCCC CTATGTTTTT GCCATTTTCA 2760

GGCTTTAACG ACCATAAGTA ATTCATCAGT TGATAAAGAT TTTGCGAATG AAATTTTTCT 2820

GTTCCCATTC GTCGTGAAAA AATGCTCTTA TAGTATTTGG CGTCAACGAT AAGTATTTTT 2880

TCTGATGAGC GAATGGTGAT GTCAGTTTCC ATTCGAGGTA ACAAATTAAG TGAATGATCC 2940

GATATACTCG ATGCATCCCA TTTTAAATAA GAGCGGGTTG TGTTTGCAGA CGTTAATTCA 3000

CGACGGCAAA ATTCATAAAG AAATTTTGA TAAAGTAATG ACATCTCTTT TTCGTTTCTT 3060

TCAAAATCAT AGAAACGGTA GTGTCCTTTG TTTTGACCTG GAATAGAATT ATTGACGATG 3120

AATTTGCAGA CACTGATAAC GAATTTATAA TAACGCGTAT TTTTCCGCC ATTGAGATAG 3180

CTGAAATGCT GCGGAGTTAA ATGAAGAGTG CTAATGCCCG GTAATTTTCT ATAAAGTGAA 3240

CGAGCTTCAT CTCTGATAGT TGAATTTAAC TTTTCATGCT TAATTAATAT GGCTAATGTG 3300

CTTTTATAAA	TTCGGTTAGC	CAGCGTGTCT	TCATTAAGCA	TATCAAAAAGT	ACTGACGGTT	3360
TTCCCATGAT	TAAGATGGAA	GCCGCGTATT	GTTTTAGCAA	ACTCTATTTCG	CCCTTTGATG	3420
CCAGGAATGA	TCTCGGTGTT	AGGATTGTAA	TCAAGCTCAA	GCCCTCGGCG	TGAAAAGCTGT	3480
AAAACCCCTT	TATTTAATAC	ATACCCCAGG	ATATCAAGAA	GATTGTTACC	GGGTATGGCT	3540
TCAAGGTTTG	CCTGCTTAAT	TTCCTGTAAA	TAACCCCATG	CATAGGTAAG	CATGTAATAG	3600
ATATTACGGA	CAGGTATCAC	GGGCTGTTCC	ACTATGAGTC	CCCTAATAAT	TTGTTGGTCC	3660
ATTTCTGTTG	TTTATAGGGG	TCATCAAAGA	AATATTCTTC	GAGTAAAGGG	GCGATATCCG	3720
TCATCACAAT	TTCATTAAGC	CATTGCGTAT	CCGAGAGAGT	GCCATCTTCC	AACCCACAGC	3780
AGAAGTAACT	ATGCCCAATG	CGGAATCCTT	TCCCAAGGAT	AGTGGCCTCT	TTGCTGATTT	3840
CCTGGTTCAA	CTCGTTCATT	TTTTGGCATA	AAGACTCAAC	AAATGAAGGT	TCTGCTTTTT	3900
TATTCAGTAA	AAAATTCCGG	AACTGTGGTG	TATCAAAACC	TGGCTCAATA	TCTATGAAAG	3960
AAAATCGTCT	GCGTAGGGCA	TAGTCAACAA	CGGCCAGAGA	GCGATCGGCA	GTATTCATTA	4020
AACCGATGAT	ATAAACATTTC	TCCGGGACAT	AGAATCGTTC	TTCATCGTTT	TCGGAGTAGG	4080
TTAGGGGAAC	AGACCAGTTT	TCACCTCGTT	TATCATGTTC	CATTAACATC	ATCACTTCGC	4140
CAAATACTTT	ACTGAGATTG	GCACGATTGA	TTTCATCTAT	AATAAAAATA	TACTTTTTCT	4200
CTGGCTGCTC	TTTAGCTTGC	TGACAAAAAT	TGTAAATAT	GCCGTCTTTA	CGTCGGAAGC	4260
CGACGCCATT	CGGACGATAG	CCCTGTATAA	AATCCTCATA	GCTATAAGAT	TGATGGAACT	4320
GAACCATATT	GACGCGTTGC	GGAGCCTTTT	CTCCTGTCAG	CAAGTAAGCC	AGACGGCGTG	4380
CAACAAAGGT	TTTTCCAACG	CCGGGCGGCC	CCTGGAGGAT	AATATTTTTT	TTGATGGTTA	4440
ATCGTTTGAG	TATCGTCTCT	ATTGTGGTTT	CAGGGATAAA	CAAATCATTT	AACGCATCTT	4500
CCAGACAGTA	TGATTCAGTT	TTTGACATAG	GTGGAATAAC	ACTCTTGCCA	GAATTAAATA	4560
TTAATTTATA	GTCGTTGATT	ATGTTGTCCA	GCATAGAGGC	AAATCGGGTG	TAATCAATAC	4620
CCTGTGAGAC	TTTTTGGGAA	CAGGCGTAAT	AGGACTGTCC	GTATTTTTTA	GGATATACAC	4680
CCGAAGTTGC	CTGAAAATAC	TCTGCGATTG	TTTTAGGTAT	GTCTGAAGAG	AACTGCCATT	4740
GGGCATGTGG	TTCATTCTGT	TCGCTTATAC	CATAAGCCAA	AACCAACTCA	TCAAAATCTT	4800
TATAATAGAG	AATAACGGGA	TATATACCGT	TAGAAGCTTC	CTGACCTTCT	CCAAGAAATG	4860
CAAACCAGGG	AATAGACGTA	AAATTACCAT	AACCGAAACT	CAATTTTACT	CGCAGGTTAC	4920
GGTAAGACGT	TGGATAATCT	TTAGTGGAAT	GCGAACGTTG	TTGCTGTGCT	TGCTTAATAA	4980
ATTTTTCAAT	CCAGGGTTGA	ATAGATTCCA	TAAGATATGC	CTTCCTCATT	GCTAAGCCTC	5040



TATTATCGCT TTCGCAACGT ACTGAAACAA TAGATTTTTTA CTGCAAAATC AGACTGGTAA 5100  
ATATTTACTG AGGGGGAAAG TTTCTATTGA GTCAGTGGA GGCTCCCGGT GGTAAACCGG 5160  
GAGTAAACGC TGTTACGCGA CTTTCTGTTT ACCGGCAATC ACTCCAATAA ACGCCTGCAC 5220  
CTGCTTTTGT TTACGCGCCG ACAGTTTGCA CACCTGGCGT AGCGACTGCA TCAGTTCGCT 5280  
CTCCTCGGCG GCGGGTGGTT GGGCGGTGAG GACAATACAG CCTTCCATCA CTTTGACATC 5340  
TACCGCCGTG CCAGTGGA AACC GGCGG CTTCCAGCCAC TGACCTTTCA GGGTGATGGC 5400  
GGGAATACGG CTGTAATCCG GGTAGCGACT CGCATAACCG ACGGTGACAT GACGGTTATT 5460  
TGCCGGGGAG ACTTCTGCTT CGAACGGTTG TGCAATAGAA TGCGTGTGAG TCATAACTGC 5520  
TATTCTCCAG GAATAGTGAT TGTGATTAGC GATGCGGGTG TGTGGCGCA CATCCGCACC 5580  
GCGCTAAATA CCTGTATATA TCATCAGTAA ATATGGGGAA AGTCCAGCTA AAAATAGAAT 5640  
AAAATGGGCA ATTTCTGGAA TGATTTAAAT ATATTTATGT GGGTTATGAT TGGCGTGAAA 5700  
TAATAAAAAG CGCACCGGAA AGGTGCGCCA GAAAATAATG TTCAGGATTT TTTACGTGAG 5760  
GCTTTTTTAC CCCCCTAGC TGCGCGTTCA GCTTTGATTT TTTCCAGCAA CGCGGCGGCG 5820  
CTGTTTTCTC CGCTGATCAA ATCCGGGTTT TCGGCCCCGCC ACTGGGCGGT AAGTTCACCA 5880  
CGGAACGCTT TTGCCAGGAT GGATTGCGTC AGGTTGTTGA CGCGGGCTAA GCGGTTGTTG 5940  
ACCTGTTTTT CTATGGTGTC GCGTAGGCG AAGAGTTGCT CGACGCGGCG AACGATTTCTG 6000  
GCTTGTTCTT TTA CTGGAGG TAATAAAACA ACTTGGGATT TGATATCTTT TCCTGAAATA 6060  
CCTTTTTGAC CAGAAGTTGT TTTACGCAG TTCATCATTG CATTTCGTGC TGAGGGGGAT 6120  
GAAAAAATA TTTCGATATA TTCTGGTAAA GCATCTTTGG TTAATCGAGC TCGAATAAGT 6180  
TTATCAGGAT ATAGCAAATT TTGATGTTGT AATTTTTTCA ATAACCCACA AACACCAACA 6240  
AATTCTAAAC TTCCGTTATA GCGAGTAAAT AAAAGATCTC CATCTTGTA TTTGTGGCGG 6300  
TTTAGTTCAC TTTCTGAACA TTCTAGAGTC GACCTGCAGG CATGCAAGCT TGGCGTAATC 6360  
ATGGTCATAG CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC ACAACATACG 6420  
AGCCGGAAGC ATAAAGTGTA AAGCCTGGGG TGCCTAATGA GTGAGCTAAC TCACATTAAT 6480  
TGCGTTGCGC TCACTGCCCC CTTTCCAGTC GGGAAACCTG TCGTGCCAGC GGATCCTCTA 6540  
CGCCGGACGC ATCGTGCCG GCATCACCGG CGCCACAGGT GCGGTTGCTG GCGCCTATAT 6600  
CGCCGACATC ACCGATGGGG AAGATCGGGC TCGCCACTTC GGGCTCATGA GCGCTTGTTT 6660  
CGGCGTGGGT ATGGTGGCAG GCCCCGTGGC CGGGGGACTG TTGGGCGCCA TCTCCTTGCA 6720  
TGCACCATTC CTTGCGGCGG CGGTGCTCAA CGGCCTCAAC CTACTACTGG GCTGCTTCCT 6780

AATGCAGGAG TCGCATAAGG GAGAGCGTCG ACCGATGCCC TTGAGAGCCT TCAACCCAGT 6840  
CAGCTCCTTC CGGTGGGCGC GGGGCATGAC TATCGTCGCC GCACTTATGA CTGTCTTCTT 6900  
TATCATGCAA CTCGTAGGAC AGGTGCCGGC AGCGCTCTGG GTCATTTTCG GCGAGGACCG 6960  
CTTTCGCTGG AGCGCGACGA TGATCGGCCT GTCGCTTGCG GTATTTCGAA TCTTGCACGC 7020  
CCTCGCTCAA GCCTTCGTCA CTGGTCCCGC CACCAAACGT TTCGGCGAGA AGCAGGCCAT 7080  
TATCGCCGGC ATGGCGGCCG ACGCGCTGGG CTACGTCTTG CTGGCGTTTCG CGACGCGAGG 7140  
CTGGATGGCC TTCCCCATTA TGATTCTTCT CGCTTCCGGC GGCATCGGGA TGCCCGCGTT 7200  
GCAGGCCATG CTGTCCAGGC AGGTAGATGA CGACCATCAG GGACAGCTTC AAGGATCGCT 7260  
CGCGGCTCTT ACCAGCCTAA CTTCGATCAT TGGACCGCTG ATCGTCACGG CGATTTATGC 7320  
CGCCTCGGCG AGCACATGGA ACGGGTTGGC ATGGATTGTA GGCGCCGCC TATACCTTGT 7380  
CTGCCTCCCC GCGTTGCGTC GCGGTGCATG GAGCCGGGCC ACCTCGACCT GAATGGAAGC 7440  
CGGCGGCACC TCGCTAACGG ATTCACCACT CCAAGAATTG GAGCCAATCA ATTCTTGCGG 7500  
AGAACTGTGA ATGCGCAAAC CAACCCTTGG CAGAACATAT CCATCGCGTC CGCCATCTCC 7560  
AGCAGCCGCA CGCGGCGCAT CTCGGGCAGC GTTGGGTCCT GGCCACGGGT GCGCATGATC 7620  
GTGCTCCTGT CGTTGAGGAC CCGGCTAGGC TGGCGGGGTT GCCTTACTGG TTAGCAGAAT 7680  
GAATCACCGA TACGCGAGCG AACGTGAAGC GACTGCTGCT GCAAAACGTC TGCGACCTGA 7740  
GCAACAACAT GAATGGTCTT CGGTTTCCGT GTTTCGTAAA GTCTGGAAAC GCGGAAGTCC 7800  
CCTACGTGCT GCTGAAGTTG CCCGCAACAG AGAGTGGAAC CAACCGGTGA TACCACGATA 7860  
CTATGACTGA GAGTCAACGC CATGAGCGGC CTCATTTCTT ATTCTGAGTT ACAACAGTCC 7920  
GCACCGCTGC CGGTAGCTCC TTCCGGTGGG CGCGGGGCAT GACTATCGTC GCCGCACTTA 7980  
TGA CTGTCTT CTTTATCATG CAACTCGTAG GACAGGTGCC GGCAGCGCCC AACAGTCCCC 8040  
CGGCCACGGG GCCTGCCACC ATACCCACGC CGAAACAAGC GCCCTGCACC ATTATGTTCC 8100  
GGATCTGCAT CGCAGGATGC TGCTGGCTAC CCTGTGGAAC ACCTACATCT GTATTAACGA 8160  
AGCGCTAACC GTTTTTATCA GGCTCTGGGA GGCAGAATAA ATGATCATAT CGTCAATTAT 8220  
TACCTCCACG GGGAGAGCCT GAGCAAACCTG GCCTCAGGCA TTTGAGAAGC ACACGGTCAC 8280  
ACTGCTTCCG GTAGTCAATA AACCGGTAAA CCAGCAATAG ACATAAGCGG CTATTTAACG 8340  
ACCCTGCCCT GAACCGACGA CCGGGTCGAA TTTGCTTTTCG AATTTCTGCC ATTCATCCGC 8400  
TTATTATCAC TTATTCAGGC GTAGCAACCA GGCGTTTAAG GGCACCAATA ACTGCCTTAA 8460  
AAAAATTACG CCCCGCCCTG CCACTCATCG CAGTACTGTT GTAATTCATT AAGCATTCTG 8520

CCGACATGGA AGCCATCACA GACGGCATGA TGAACCTGAA TCGCCAGCGG CATCAGCACC 8580  
 TTGTCGCCTT GCGTATAATA TTTGCCCATG GTGAAAAACGG GGGCGAAGAA GTTGTCCATA 8640  
 TTGGCCACGT TTAAATCAAA ACTGGTGAAA CTCACCCAGG GATTGGCTGA GACGAAAAAC 8700  
 ATATTCTCAA TAAACCCTTT AGGGAAATAG GCCAGGTTTT CACCGTAACA CGCCACATCT 8760  
 TCGGAATATA TGTGTAGAAA CTGCCGGAAA TCGTCGTGGT ATTCACTCCA GAGCGATGAA 8820  
 AACGTTTCAG TTTGCTCATG GAAAACGGTG TAACAAGGGT GAACACTATC CCATATCACC 8880  
 AGCTCACCGT CTTTCATTGC CATACG 8906

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3190 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "pRM2 (target plasmid)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCGCCCAATA CGAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA 60  
 CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT 120  
 CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT 180  
 TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG AATTCGAGCT 240  
 CGGTACCCGG GGATCCTCTA GAGTCGAGAT GCCGCATGTG GAAGAGGTGA TTGCACCGAT 300  
 CTTCTACACC GTTCCGCTGC AGCTGCTGGC TTACCATGTC GCGCTGATCA AAGGCACCGA 360  
 CGTTGACCAG CCGCGTAACC TGGCAAAATC GGTTACGGTT GAGTAATAAA TGGATGCCCT 420  
 GCGTAAGCGG GGCATTTTTC TTCCTGTTAT GTTTTTAATC AAACATCCTG CCAACTCCAT 480  
 GTGACAAACC GTCATCTTCG GCTACTTTTT CTCTGTCACA GAATGAAAAT TTTCTGTCAT 540  
 CTCTTCGTTA TTAATGTTG TAATTGACTG AATATCAACG CTTATTTAAA TCAGACTGAA 600  
 GACTTATCTC TCTCTGTCAT AAAACTGTCA TATTCCTTAC ATATAACTGT CACCTGTTTG 660  
 TCCTATTTTG CTTGTCGTAG CCAACAAACA ATGCTTTATG AATCCTCCCA GGAGACATTA 720  
 TGAAAGTTAT GCGTACCACC GTCGCAACTG TTGTCGCCGC GACCTTATCG ACCTGCAGGC 780  
 ATGCAAGCTT GGCACCTGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA 840

CCCAACTTAA	TCGCCTTGCA	GCACATCCCC	CTTTCGCCAG	CTGGCGTAAT	AGCGAAGAGG	900
CCCGCACCGA	TCGCCCTTCC	CAACAGTTGC	GCAGCCTGAA	TGGCGAATGG	CGCCTGATGC	960
GGTATTTTCT	CCTTACGCAT	CTGTGCGGTA	TTTCACACCG	CATATGGTGC	ACTCTCAGTA	1020
CAATCTGCTC	TGATGCCGCA	TAGTTAAGCC	AGCCCCGACA	CCCGCCAACA	CCCCTGACG	1080
CGCCCTGACG	GGCTTGCTCTG	CTCCCGGCAT	CCGCTTACAG	ACAAGCTGTG	ACCGTCTCCG	1140
GGAGCTGCAT	GTGTCAGAGG	TTTTCACCGT	CATCACCAGAA	ACGCGCGAGA	CGAAAGGGCC	1200
TCGTGATACG	CCTATTTTTTA	TAGGTTAATG	TCATGATAAT	AATGGTTTCT	TAGACGTCAG	1260
GTGGCACTTT	TCGGGGAAAT	GTGCGCGGAA	CCCCTATTTG	TTTATTTTTTC	TAAATACATT	1320
CAAATATGTA	TCCGCTCATG	AGACAATAAC	CCTGATAAAT	GCTTCAATAA	TATTGAAAAA	1380
GGAAGAGTAT	GAGTATTCAA	CATTTCCGTG	TCGCCCTTAT	TCCCTTTTTT	GCGGCATTTT	1440
GCCTTCCTGT	TTTTGCTCAC	CCAGAAACGC	TGGTGAAAGT	AAAAGATGCT	GAAGATCAGT	1500
TGGGTGCACG	AGTGGGTTAC	ATCGAACTGG	ATCTCAACAG	CGGTAAGATC	CTTGAGAGTT	1560
TTCGCCCCGA	AGAACGTTTT	CCAATGATGA	GCACTTTTAA	AGTTCTGCTA	TGTGGCGCGG	1620
TATTATCCCG	TATTGACGCC	GGGCAAGAGC	AACTCGGTCTG	CCGCATACAC	TATTCTCAGA	1680
ATGACTTGGT	TGAGTACTCA	CCAGTCACAG	AAAAGCATCT	TACGGATGGC	ATGACAGTAA	1740
GAGAATTATG	CAGTGCTGCC	ATAACCATGA	GTGATAACAC	TCCGGCCAAC	TTACTTCTGA	1800
CAACGATCGG	AGGACCGAAG	GAGCTAACCG	CTTTTTTGCA	CAACATGGGG	GATCATGTAA	1860
CTCGCCTTGA	TCGTTGGGAA	CCGGAGCTGA	ATGAAGCCAT	ACCAAACGAC	GAGCGTGACA	1920
CCACGATGCC	TGTAGCAATG	GCAACAACGT	TGCGCAAAC	ATTAAC	TGGC	1980
CTCTAGCTTC	CCGGCAACAA	TTAATAGACT	GGATGGAGGC	GGATAAAGTT	GCAGGACCAC	2040
TTCTGCGCTC	GGCCCTTCCG	GCTGGCTGGT	TTATTGCTGA	TAAATCTGGA	GCCGGTGAGC	2100
GTGGGTCTCG	CGGTATCATT	GCAGCACTGG	GGCCAGATGG	TAAGCCCTCC	CGTATCGTAG	2160
TTATCTACAC	GACGGGGAGT	CAGGCAACTA	TGGATGAACG	AAATAGACAG	ATCGCTGAGA	2220
TAGGTGCCTC	ACTGATTAAG	CATTGGTAAC	TGTCAGACCA	AGTTTACTCA	TATATACTTT	2280
AGATTGATTT	AAAAC	TTCAT	TTTTAATTTA	AAAGGATCTA	GGTGAAGATC	2340
ATCTCATGAC	CAAAATCCCT	TAACGTGAGT	TTTCGTTCCA	CTGAGCGTCA	GACCCCGTAG	2400
AAAAGATCAA	AGGATCTTCT	TGAGATCCTT	TTTTTCTGCG	CGTAATCTGC	TGCTTGCAAA	2460
CAAAAAAACC	ACCGCTACCA	GCGGTGGTTT	GTTTGCCGGA	TCAAGAGCTA	CCAACTCTTT	2520
TTCCGAAGGT	AACTGGCTTC	AGCAGAGCGC	AGATACCAA	TACTGTCCTT	CTAGTGTAGC	2580

CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA 2640  
 TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA 2700  
 GACGATAGTT ACCCGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTTCG TGCACACAGC 2760  
 CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA 2820  
 GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA 2880  
 CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCTG 2940  
 GGTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC 3000  
 TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG 3060  
 CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG 3120  
 AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG 3180  
 AAGCGGAAGA 3190

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotide (NLC95)  
used to analyze products of transposition."

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATAATCCTTA AAAACTCCAT TTCCACCCCT 30

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 29 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Oligonucleotide (NLC209)  
used to analyze products of transposition"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTGATTGCAC CGATCTTCTA CACCGTTCC

29

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide (NLC429)  
used to analyze products of transposition"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTTCACCGTC ATCACCGAAA CGCGCGAGAC

30

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide (NLC430)  
used to analyze products of transposition"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AATGACTTGG TTGAGTACTC ACCAGTCACA

30

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide (NLC431)  
used to analyze the products of transposition"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGAACGAAA TAGACAGATC GCTGAGATAG

30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide (NLC432)  
used to analyze products of transposition"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAAGACGATA GTTACCGGAT AAGGCGCAGC

30

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Oligonucleotide (NLC94)  
used for sequence determination"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAAGTCCAGT ATGCTTTTTC ACAGCATAAC

30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asn Tyr Asn Arg Asn  
1 5

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 5 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Asn Tyr Thr Arg Asn  
1                    5